PATTERN DISCOVERY TECHNIQUES FOR DETERMINING MAXIMAL IRREDUNDANT AND REDUNDANT MOTIFS

Abstract of the Disclosure

5   Basis motifs are determined from an input sequence through an iterative technique that begins by creating small solid motifs and continues to create larger motifs that include "don't care" characters and that can include flexible portions. The small solid motifs, including don't care characters and flexible portions, are concatenated to create larger motifs. During each iteration, motifs are trimmed to remove redundant motifs and

10   other motifs that do not meet certain criteria. The process is continued until no new motifs are determined. At this point, the basis set of motifs has been determined. The basis motifs are used to construct redundant motifs. The redundant motifs are formed by determining a number of sets for selected basis motifs. From these sets, unique intersection sets are determined. The redundant motifs are determined from the unique

15   intersection sets and the basis motifs. This process continues, by selecting additional basis motifs, until all basis motifs have been selected.